Title

A genotyping protocol for multiple tissue types from the polyploid tree species Sequoia sempervirens (Cupressaceae).

Journal

Applications in plant sciences, 3(3)

ISSN

2168-0450

Authors

Narayan, Lakshmi
Dodd, Richard S
O'Hara, Kevin L

Publication Date

2015-03-01

DOI

10.3732/apps.1400110

Peer reviewed

Applications in Plant Sciences

Applications in Plant Sciences 2015 3(3): 1400110

Application Article

A genotyping protocol for multiple tissue types from the polyploid tree species *Sequoia sempervirens* (Cupressaceae)[1]

Lakshmi Narayan[2,3], Richard S. Dodd[2], and Kevin L. O'Hara[2]

[2]Department of Environmental Science, Policy, and Management, University of California, Berkeley, 130 Mulford Hall #3114, Berkeley, California 94720-3114 USA

- *Premise of the study:* Identifying clonal lineages in asexually reproducing plants using microsatellite markers is complicated by the possibility of nonidentical genotypes from the same clonal lineage due to somatic mutations, null alleles, and scoring errors. We developed and tested a clonal identification protocol that is robust to these issues for the asexually reproducing hexaploid tree species coast redwood (*Sequoia sempervirens*).
- *Methods:* Microsatellite data from four previously published and two newly developed primers were scored using a modified protocol, and clones were identified using Bruvo genetic distances. The effectiveness of this clonal identification protocol was assessed using simulations and by genotyping a test set of paired samples of different tissue types from the same trees.
- *Results:* Data from simulations showed that our protocol allowed us to accurately identify clonal lineages. Multiple test samples from the same trees were identified correctly, although certain tissue type pairs had larger genetic distances on average.
- *Discussion:* The methods described in this paper will allow for the accurate identification of coast redwood clones, facilitating future studies of the reproductive ecology of this species. The techniques used in this paper can be applied to studies of other clonal organisms as well.

Key words: clonal; coast redwood; genotyping; null alleles; polyploidy; *Sequoia sempervirens*.

Coast redwood (*Sequoia sempervirens* (D. Don) Endl.) is an iconic species and important source of timber production and carbon storage in northern California. It is also one of a few conifer species able to produce basal sprouts as a form of natural clonal reproduction. Redwood trees commonly regenerate from cut stumps, fallen logs, or roots (Neal, 1967; Del Tredici, 1998). This vegetative reproduction may lead to the dominance of a small number of clones over a large area and the long-term persistence of genotypes. In the case of redwoods, which are extremely long-lived as individual stems, clonal reproduction could theoretically lead to the persistence of single genotypes for tens of thousands of years. Given the role of coast redwood as a valuable endemic and timber species, surprisingly little is known about the extent of clonal reproduction and patterns of genotypic diversity throughout its range.

[1]Manuscript received 13 November 2014; revision accepted 6 February 2015.

The authors thank B. T. Caldwell, D. A. Jones, K. I. McGown, J. A. Sherwood, and V. Narayan for their assistance in field data collection for this project. A. C. Ragsac, L. A. Hall, and C. M. Denby contributed to the laboratory methods. P. V. A. Fine and three reviewers contributed valuable feedback on the manuscript. California State Parks and the National Park Service provided access to our field sites. This work was funded by the National Science Foundation's Graduate Research Fellowship Program (NSF 14-590), a Redwood Research Grant from Save the Redwoods League, and a University of California Berkeley Bridging Grant. Participation by R.S.D. and K.L.O. was partially supported by the USDA National Institute of Food and Agriculture.

[3]Author for correspondence: lakshmi.narayan@berkeley.edu

doi:10.3732/apps.1400110

Previous studies of clonal patterns in old growth (Rogers, 2000; Rogers and Westfall, 2007) and second-growth (Douhovnikoff et al., 2004) coast redwood forests using allozyme markers and amplified fragment length polymorphisms, respectively, found that multiple genotypes were often intermingled, and that members of the same clone could be found up to 340 m apart. Due to the challenge of collecting foliage from the canopy of dominant redwood trees, no study to date has been able to comprehensively sample all trees in a forest area. Microsatellite markers may facilitate genetic studies of trees where high-quality foliar tissue is not available because their use requires relatively low concentrations of template DNA. Additionally, microsatellites are generally species-specific, which eliminates potential interspecific contamination in samples with low concentrations of DNA from the species of interest.

One factor that complicates genotyping coast redwoods using microsatellite markers is its hexaploid condition. In genetic analyses of polyploid organisms, it is difficult to (1) discern copy number of alleles in microsatellite scans, and (2) accurately score microsatellite scans with potentially higher numbers of alleles. For coast redwood, copy number can be determined for a homozygote (one allele) or a full heterozygote (six alleles), but for partial heterozygotes, copy number is impossible to determine with certainty. One method for polyploid organisms is to estimate copy number using the peak size on microsatellite scans (Esselink et al., 2004). However, implementing this method becomes more challenging with increasing ploidy, and not all marker sets have consistent enough amplification to confidently employ this method.

Additionally, testing the fidelity of amplification products is complicated in polyploids. Because allele copy number cannot

typically be resolved exactly, tests for null alleles and other PCR artifacts that require calculation of exact allele frequencies cannot be used on polyploid organisms (Dufresne et al., 2014). Microsatellite scans with many alleles make it more challenging to determine the presence of stutter bands (Pfeiffer et al., 2011). For coast redwood, it is possible to observe between one and six alleles in a microsatellite scan. If the size difference between alleles is within several base pairs, it can be difficult to distinguish between stutter and true alleles.

Another challenge in determining the genotypic identity of clonal plants regardless of ploidy level is the possibility of somatic mutation, where a mutation occurs that changes the genotype of an individual in a clonal lineage. For coast redwood, basal sprouting often occurs as response to disturbance, such as fire or timber harvesting (Neal, 1967; Lorimer et al., 2009, Ramage et al., 2010). Somatic mutation in basal sprouts has the potential to confound genotyping studies seeking to identify the origin of shoots, particularly in cases where different tissue types are being sampled for clonal identification.

Given that two ramets from a clonal plant may differ in their genotype due to the presence of null alleles, scoring errors, or somatic mutation, the concept of identifying clones that belong to a multilocus lineage (MLL) has been proposed to identify clonal lineages that may not be identical in genotype (Arnaud-Haond et al., 2007a). Here, we used microsatellite data to identify MLLs using multiple tissue types from coast redwood. Existing protocols were modified to extract and amplify DNA from redwood cambium, and samples of cambium and leaf tissue from the same trees were compared to ensure consistency between tissue types in our genotyping protocol. We also developed a novel protocol to improve accuracy of microsatellite scoring. Monte Carlo simulations were used to calculate probability of identity and explore the effect of null alleles on genotyping accuracy. In addition to providing genotyping methods for future studies of coast redwood, these protocols should be applicable to genotyping other polyploid species.

METHODS

***Sample collection*—**Samples were collected in square 1-ha plots in old-growth redwood forests in northern California. Plots were located in areas classified as "old-growth" on Save the Redwoods League maps where coast redwood was the dominant species. Two 1-ha plots were located at Big Basin Redwoods State Park (37.18056′N, 122.23278′W; 37.18528′N, 122.21444′W), two at Humboldt Redwoods State Park (40.34833′N, 123.92444′W; 40.3402′N, 123.94833′W), one at Redwood National Park (41.30750′N, 124.02667′W), and one at Prairie Creek Redwoods State Park (41.37250′N, 124.02528′W). All trees larger than 10 cm dbh were mapped, measured for diameter, classified by canopy position and strata (Oliver and Larson, 1996), and identified to species. All coast redwood trees were cored for cambium/sapwood samples using a 5.15-mm diameter increment borer. The increment borer was dipped in and sprayed with 10% bleach, rinsed, and dried with several lengths of clean yarn between trees. Cambium samples were preserved in bags of silica gel. Wherever foliage, epicormic sprouts, or basal sprouts (hereafter referred to collectively as "leaf" samples) were accessible, they were collected in resealable storage bags with a few drops of distilled water. All samples were stored in a 4°C freezer within two weeks of collection.

***DNA extraction*—**Leaf samples were cut and ground for 1 min in a Mini Beadbeater (BioSpec Products, Bartlesville, Oklahoma, USA) using a combination of 2.5-mm and 6.35-mm glass beads in XXTuff Reinforced 2-mL Microvials (BioSpec Products). Cambium samples were freeze dried for at least 72 h using a FreeZone 12 Freeze Dry System (Labconco, Kansas City, Missouri, USA) then ground to a powder in XXTuff Reinforced 2-mL Microvials using 6.35-mm chrome-steel beads (BioSpec Products). Cambium samples were ground in three 1-min intervals. Between grinding intervals, samples were placed on ice for 5 min to prevent degradation from overheating. DNA was extracted from both leaf and cambium samples using a modified cetyltrimethylammonium bromide (CTAB) method (Cullings, 1992).

***Primer development*—**We tested primers that were developed from genomic libraries by Bruno and Brinegar (2004) and Douhovnikoff and Dodd (2011) for use in this study. To test primers, we used a set of 21 samples from Humboldt Redwoods State Park (HRSP) and a control tree from the University of California, Berkeley (UC Berkeley), campus. Samples from HRSP were in sets of three that included foliage, epicormic, and basal samples from the same tree. We initially screened primers by amplifying fragments from our test samples and visualizing the product using gel electrophoresis. If a primer amplified fragments showing consistency within trees and polymorphism between trees, we ran PCRs with fluorescent-labeled primers with different salt concentrations and temperature cycling protocols to see which were polymorphic and amplified well. We found that primers SEQ8E8 (dinucleotide repeats) and SEQ18D7-3 (trinucleotide repeats) from Bruno and Brinegar (2004) and RW28 and RW39 (tetranucleotide repeats) from Douhovnikoff and Dodd (2011) amplified well and were polymorphic. In addition to the four previously developed primers, we also developed two new primers, RW56 and RWDI11. Cloning and sequencing followed Douhovnikoff and Dodd (2011). From these sequences, we developed an additional primer for a tetranucleotide repeat region (RW56) and an additional primer for a dinucleotide repeat region (RWDI11).

***PCR optimization*—**For all primers, we optimized amplification by testing magnesium chloride ($MgCl_2$) concentrations between 1.5 and 3.0 mM and by modifying the number of cycles and range of annealing temperatures in the thermocycling protocols. The optimal salt concentration for all primers was 3.0 mM $MgCl_2$. Optimal annealing temperature range and number of cycles differed between primers, but all protocols were touchdown protocols that consisted of: (1) an initial denaturing period of 3 min at 94°C; (2) 27–35 cycles of denaturing for 1 min at 94°C, 1 min of annealing, where the annealing temperature was lowered each cycle, and 1 min of extension at 72°C; (3) one cycle of denaturing for 1 min at 94°C, 1 min of annealing at 45°C, and 1 min of extension at 72°C; and (4) a final extension at 72°C for 2 min. Number of cycles and annealing temperatures for each primer are given in Table 1. PCRs took place in 10-μL volumes consisting of 1 μL of 1 : 10 diluted template DNA, 1× PCR Buffer (Invitrogen Life Technologies, Carlsbad, California, USA), 3.0 mM $MgCl_2$ (Invitrogen Life Technologies), 800 μM dNTPs, 0.6 μM each forward and reverse primers (Integrated DNA Technologies, Coralville, Iowa, USA), 0.25 μg/μL bovine serum albumin (New England Biolabs, Ipswich, Massachusetts, USA), 0.25 units *Taq* Polymerase (Invitrogen Life Technologies), and water to bring the final volume to 10 μL. Forward primers were labeled with either 6-FAM or HEX fluorescent dyes (Table 1). For the marker SEQ18D7-3, the reverse primer was labeled instead of the forward primer. Leaf and cambium PCRs were always separate to prevent contamination of cambium samples, which could potentially have a lower concentration of template DNA due to fewer living cells in woody tissue than in leaf tissue. PCR product was diluted 1 : 10, and fragments were analyzed with GeneScan 500 LIZ Size Standard (Applied Biosystems, Foster City, California, USA) on an ABI 3730 DNA Analyzer (Applied Biosystems) at the Evolutionary Genetics Laboratory in the Museum of Vertebrate Zoology at UC Berkeley. A positive control sample from a tree on the UC Berkeley campus and a blank were included on each plate.

***Allele scoring*—**Microsatellite data were analyzed with GeneMapper v4.0 software (Applied Biosystems). To make our allele scoring protocol more robust against the accidental scoring of stutter peaks or noise, we created bins only for alleles that amplified in at least two different tissue types from the same tree. For example, if we found a new allele in a cambium sample, we extracted and amplified a second sample from an alternate tissue type (foliage, epicormic, or basal) collected from the same tree to verify the allele. If an allele did not amplify in multiple tissue types, a bin was not added for that allele. We used the GeneMapper software to score alleles and manually checked and rescored samples as necessary.

Given the quality of our primers and our comparison of different tissue types, we did not think that we would be able to accurately estimate copy number in partial heterozygotes from allele peak size as described in Esselink et al. (2004). Instead, alleles were recorded as either present or absent in each sample.

***Clonal assignment protocol*—**To determine which trees were part of the same MLL, we used a protocol described by Arnaud-Haond et al. (2007b). We

TABLE 1. Information on previously published and newly developed microsatellite primers for *Sequoia sempervirens*.

Locus (label)	GenBank accession no.	Primer sequences (5′–3′)	No. of PCR extension cycles	T_a (°C)	Allele size range (bp)	A[a]	H_o[a]	No. of individuals with 1/2/3/4/5/6 alleles detected[a]
RW28 (FAM)	GU969047	F: GATAGATAAATAGATGGATAG R: TTTTTAAGGTTTCATGGATAAGTACAA	35	65–50	187–342	19	0.39	219/106/29/5/0/0
RW39 (FAM)	GU969046	F: CCATAAGGTTGAAATGAAGAAAAA R: GTTGATTGATCGTTGGTTGG	30	65–50	240–470	69	1.00	2/9/52/138/149/97
SEQ18D7-3 (HEX)	AY562168	F: GCAAAAAGGGAATTGTAATTGGGTTCA R: CCCTAGGTCTAGGCTACGCGACTTG	27	67–52	124–183	13	0.68	136/180/88/15/5/0
SEQ8E8 (FAM)	AY562169	F: ATACTCACCCTTACACGGGC R: AAATGCCTTGATGAAGCAAAA	28	67–52	112–185	21	0.24	246/65/9/2/0/0
RW56 (HEX)	KP055095	F: CTTGACATCATCCATAGCT R: AAATTGCAAGGGGTGCAA	30	69–54	189–259	18	0.89	51/116/166/89/23/1
RWDI11 (HEX)	KP055096	F: GGACCAAATGCCCTGAAC R: GCCAAGCCATATGGGTTTG	30	63–48	215–268	29	1.0	1/19/65/174/142/48

Note: A = number of alleles; H_o = observed heterozygosity; T_a = annealing temperature.
[a] A, H_o, and numbers of individuals with each possible allele count were calculated using combined data from all six study plots.

calculated the pairwise genetic distances between samples at each site (Big Basin Redwood State Park, Humboldt Redwoods State Park, and Redwood National Park/Prairie Creek Redwoods State Park) using POLYSAT (version 1.3-2–1.3-3; Clark and Jasieniuk, 2011) in R (version 3.1.1; R Core Team, 2014). We used both the Bruvo distance metric (Bruvo et al., 2004), which takes into consideration that alleles similar in size could be closely related by mutation, and the Lynch distance metric (Lynch, 1990), which is a simpler band-sharing measure. As results from both metrics were very similar, from here forward we present results using the Bruvo metric. For a nonclonal organism with random mating, we would expect a histogram of the pairwise genetic distances between individuals to show a roughly normal distribution. For a clonal organism, we would instead expect the histogram of pairwise genetic distances to have a bimodal distribution, with one peak centered on the mean genetic distance between nonclonal individuals, and a second peak very close to zero, consisting of pairwise genetic distances between samples from the same MLL. If the genotypes of all clonal pairs are perfectly identical, the genetic distance between these samples should be zero. However, due to scoring errors, null alleles, and somatic mutation, the genetic distance between clones may be greater than zero. We planned to set the genetic distance threshold for clonal assignment at the anticipated trough between the clonal and nonclonal peaks in histograms of pairwise genetic distances.

***Probability of identity calculation*—**To calculate the probability of identity (P_{ID}), we used Monte Carlo simulations to determine the probability of drawing two indistinguishable genotypes given the overall allele frequencies from sampled individuals. Calculating P_{ID} for a polyploid is complicated by copy number ambiguity, because many allelic configurations are possible for partially heterozygous allelic phenotypes. Instead of calculating P_{ID} based on the presence or absence of alleles, we developed a protocol to account for the multiple different genotypes that could result in an identical allelic phenotype.

We used the "round-robin" method developed by Parks and Werth (1993) to calculate allele frequencies in populations of clonal plants. To calculate the allele frequency for a given primer, clonal identity of each individual was determined without data from the marker for which allele frequencies were being calculated. The data set was then trimmed to include one individual per clone, and allele frequencies were calculated using the remaining individuals. This process was repeated for each marker. Because allele frequencies could not be calculated exactly due to uncertainty in allele copy number, we used the simple allele frequency estimator in POLYSAT. The use of the simple allele frequency estimator assumes that, in a partially heterozygous sample, all alleles have an equal probability of being present in multiple copies. This estimator did not allow us to account for inbreeding or departures from Hardy–Weinberg equilibrium, which are likely given coast redwood's clonality and noncontinuous geographic distribution. However, given the complexity of accounting for these factors in polyploid organisms, we chose to use a simple allele frequency estimator that did not require us to make any assumptions about the evolution of polyploidy in coast redwood or levels of selfing in this species.

Initially, we attempted to calculate probability of identity by a "brute force" method where we first created a matrix of every possible genotypic permutation. Next, we added a column to describe the allelic phenotype of each genotype. For example, a genotype with alleles *aaabbb* would have an allelic phenotype of *ab*. Then, we summed the probability of all permutations that would yield a given allelic phenotype. For a hexaploid, this would mean that we summed the probability of all 62 genotypes that yielded the allelic phenotype *ab*. Finally, we summed the squared probabilities of each allelic phenotype to find the probability that an identical allelic phenotype would appear in two successive draws.

Unfortunately, this brute force method resulted in extremely large matrices of possible allelic configurations. For our most diverse locus, which had 69 alleles, there were greater than 1×10^{11} permutations. Instead of calculating the probability of every possible genotype, we instead used Monte Carlo simulations to approximate the probability of drawing an identical allelic phenotype twice for a given locus. To do this, we simulated single-locus genotypes for 100,000 pairs of trees based on our allele frequencies. We then assigned the appropriate allelic phenotype to each tree, and counted the number of times out of 100,000 that the paired trees had a matching allelic phenotype to estimate P_{ID} for each locus. To find the overall probability of identity, we multiplied the P_{ID} estimates from each locus, then multiplied that number by 32,942, the number of comparisons in the plot with the maximum number of sampled redwood trees (182). We were able to verify the accuracy of this method by comparing results of the Monte Carlo simulations to our brute force results from our least diverse primer, SEQ18D7-3.

***Null allele trials*—**While P_{ID} calculations gave the probability of finding an exact match in our genotyping data between sexually reproduced samples, our clonal identification protocol allowed individuals with slightly different allelic phenotypes to be assigned to the same MLL. To test the sensitivity of our genotyping protocol to null alleles, we created simulated data sets with increasing numbers of missing alleles to see how this impacted the probability of assigning sexually generated genotypes to the same MLL. In each simulation, genotypes of 182 trees were randomly generated using allele frequencies from the original data as the probability of sampling each allele. Alleles present more than once in an individual were deleted to reduce the genotype data down to allelic phenotype data, to match the allele copy number ambiguity present in the original data. Next, null alleles were deleted from individuals in roughly the same number from each marker so that each marker had an allele deleted in 30 to 31 individuals. Within each marker, alleles were deleted randomly with equal probability. If a marker only had one allele present during a round of deletions, it would be skipped, and its single allele would not be deleted, because in the actual data collection microsatellite scans that showed no alleles were rerun. Once a data set had been simulated and alleles deleted, we used the same clonal assignment protocol that was used on the original data, and determined whether any individuals had been classified into the same MLL. We simulated 100 data sets of 182 trees for each number of rounds of deletions (0–30), and counted (1) the number of simulations out of 100 that had false positives and (2) the total number of false positives present in all 100 simulations. We also calculated the average number of deletions per tree for each number of rounds of deletions, because deletions were skipped for markers with only one allele present.

***Test samples*—**To check our genotyping protocol, we tested it on 88 sets of paired samples of different tissue types from the same trees. Of these 88 sample

sets, we had similar numbers of comparisons between foliage-epicormic samples, foliage-basal samples, epicormic-basal samples, and cambium-basal samples, which allowed us to compare the average genetic distance between different tissue type pairs using an analysis of variance (ANOVA). We also assessed the effect of variation between amplification plates on genetic distance between paired samples. Pairwise genetic distance between duplicate samples was regressed on proportion of loci amplified on the same plate using a linear model. Because pairs consisting of cambium and leaf samples were always run on separate plates, we excluded them from this analysis to eliminate the potentially confounding effect of tissue type on our assessment of whether variation between plates affected genetic distances between duplicate samples. Prior to these analyses, the data set was checked for outliers and any outliers were removed from the analysis.

RESULTS

***Pairwise genetic distances*—**We plotted histograms of pairwise genetic distances for each 1-ha plot, which generally showed one peak around 0.6 and a second peak close to zero. The second peak likely resulted from scoring errors or somatic mutations causing slight variation between the genotype of clones (Fig. 1). The histograms consistently showed a trough around 0.2, so we set this as our genetic distance cutoff for clonal assignment. Using this criterion, 449 clones were identified in the 770 trees genotyped.

***Probability of identity calculation*—**We compared the estimate of P_{ID} generated from Monte Carlo simulations to our "brute force" calculation from our least diverse primer, SEQ18D7-3, which had 13 alleles, and found that 100,000 simulations were enough to give us an estimate that was accurate within 10^{-3}. When we calculated the P_{ID} from all six primers, the product, or overall P_{ID}, was less than 1.1×10^{-18}. Correcting for the number of comparisons being made in the plot with the most trees resulted in a $P_{ID} < 3.6 \times 10^{-14}$.

***Null allele trials*—**Our null allele trials showed one or fewer false positives in sets of 100 simulations up to 20 rounds of deletions, an average of 18 actual deletions (Fig. 2). When we deleted 20 rounds of alleles from allelic phenotypes in our simulations, three out of 100 simulations contained one false positive, giving an error rate of 0.03. In further simulations with increasing numbers of alleles deleted, both the number of simulations out of 100 that had false positives and the total number of false positives present in all 100 simulations continued to increase.

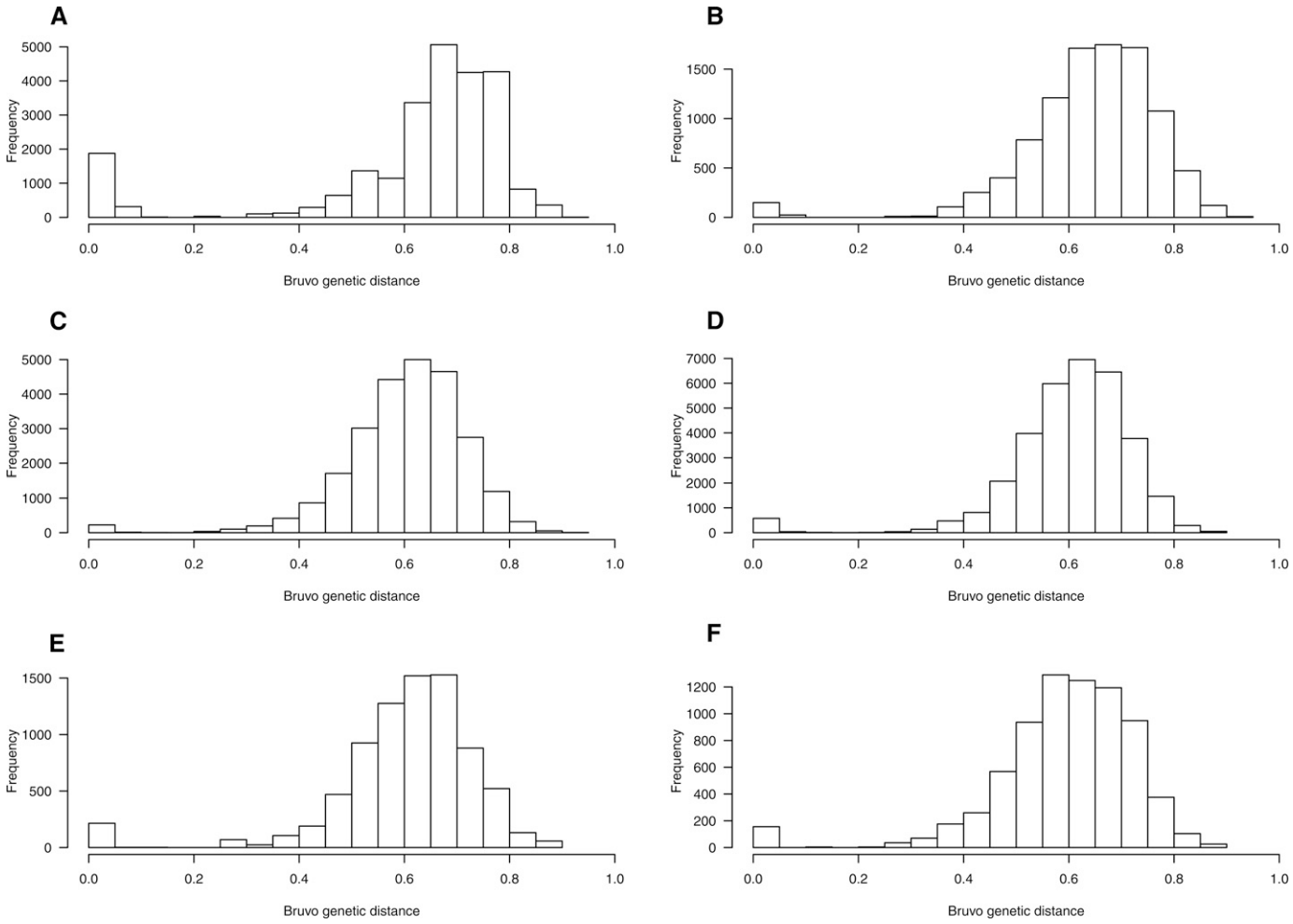

Fig. 1. Plot of pairwise Bruvo distances for (A) Big Basin Redwoods State Park 1, (B) Big Basin Redwoods State Park 2, (C) Humboldt Redwoods State Park 1, (D) Humboldt Redwoods State Park 2, (E) Redwood National Park, and (F) Prairie Creek Redwoods State Park.

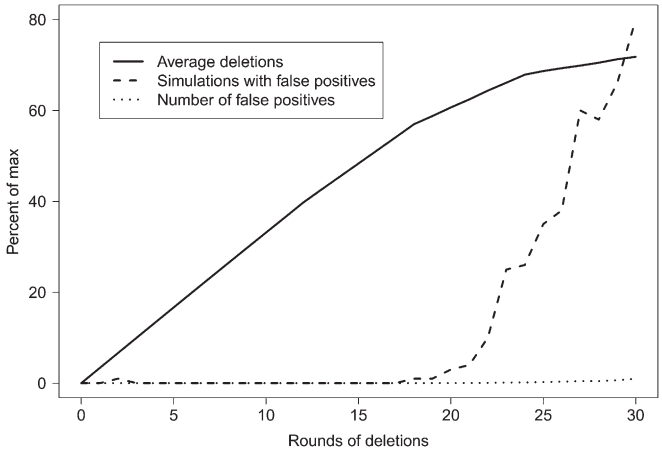

Fig. 2. Results of null allele trials. One hundred data sets were simulated at each number of rounds of deletions. Lines show average deletions, simulations with false positives, and the total number of false positives for all simulations with a given number of rounds of deletions. Results are shown as a percentage of the maximum possible value of each variable.

***Test samples*—**Of our 88 sets of paired samples, only one pair of samples from the same tree was identified as clonally distinct. This pair had a genetic distance of 0.60, and consisted of a cambium sample and a basal sample. Excluding this sample, the mean genetic distance between paired samples was 0.03 and ranged from 0 to 0.17. Forty-nine out of 87 remaining pairs had a genetic distance of zero. Of the duplicate pairs with a genetic distance greater than zero, most of these differences were due to one or two alleles being present in one sample but not the other. Eleven out of 87 pairs had alleles that were one base pair different. In all of these cases, the mismatching alleles were from the primer RWDI11, which amplified a dinucleotide repeat region.

An ANOVA comparing genetic distances between paired samples of different tissue type combinations showed a modest statistical difference between sample types ($F(3,82) = 2.93$, $P = 0.03$; Fig. 3). A Tukey's honest significant difference test showed that the genetic distance between foliage-epicormic samples was, on average, lower than cambium-basal samples ($P = 0.02$), but there were no other differences between tissue type combinations. The regression of genetic distance between duplicate samples on proportion of loci amplified on the same plate showed a small but significant negative correlation (slope = −0.040, $t(65) = -2.31$, $P = 0.02$).

DISCUSSION

Results from our probability of identity calculations, null allele simulations, and test samples suggest that our genotyping

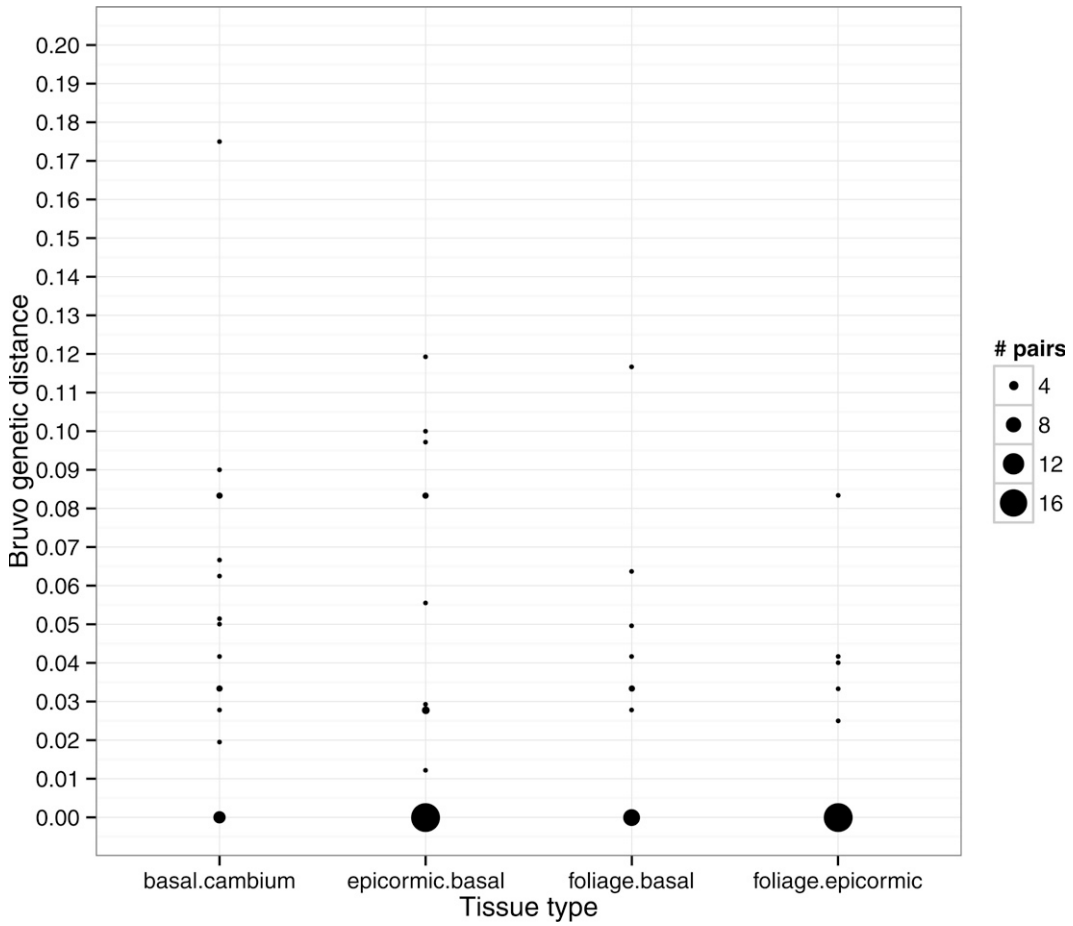

Fig. 3. Bruvo genetic distance between test samples of different tissue-type pairs. Circles are scaled to show the number of sample pairs with each genetic distance.

protocol was able to consistently identify MLLs. Optimizing PCR conditions and confirming consistent amplification of alleles before scoring allowed us to generate histograms with a consistent trough in the distribution of genetic distances between clonal and nonclonal trees. Using the genetic distance value at this trough as the threshold in our clonal assignment protocol, trees were assigned into MLLs in a way that accounted for nonzero genetic distances. Our protocol distinguished between clones collected in close physical proximity, which might be more genetically similar than individuals sampled at random from a population.

Null allele trials also suggested that our genotyping protocol was robust to the presence of null alleles. In our simulations, randomly generated allelic phenotypes were identified as clones in one in 100 or fewer simulations, with up to 18 deleted alleles. Our protocol for clonal identification may be useful for other studies of polyploid plants where null alleles are an issue, although consideration should be given to the fact that, in studies with less diverse primer sets than ours, null alleles may present more of a challenge than they do here. For both probability of identity estimation and null allele trials, we found simulations to be extremely useful. Our simulations for calculating probability of identity and investigating the robustness of our genotyping protocol could also be applied earlier in a clonal identification study to determine (1) how many markers are needed for reliable genotyping of an organism or (2) whether a highly conservative microsatellite scoring protocol that had the potential to generate null alleles would be appropriate for a given set of markers.

Results from test samples showed that our genotyping protocol was robust to the use of different tissue types. We found only one case where two samples from the same tree were not assigned to the same MLL. In this case, the samples were a basal sprout and cambium sample from the same tree, with a genetic distance of 0.6. Given this genetic distance, it seems extremely unlikely that these two samples came from the same MLL. Instead, it seems more likely that the basal and cambium samples in this pair came from different trees. During sample collection, some basal samples collected were sprouting out of the ground near the presumed parent tree, so there was some potential for misidentification. To confirm that laboratory contamination was not the reason for this mismatch, both samples were reanalyzed for all loci, but results remained the same.

Although our protocol for assignment into MLLs was robust to the use of different tissue types, different tissue type pairs varied in their average pairwise genetic differences. Pairs of duplicate samples consisting of basal and cambium tissue from the same tree had the highest average genetic distance, while foliage-epicormic pairs had the lowest. Most nonzero pairwise genetic distances between samples from the same tree were due to null alleles in one of the samples. Although it is possible that somatic mutation in the microsatellite primer regions is responsible for some missing alleles, it seems unlikely that this is responsible for the number of null alleles we observed in duplicate samples. Instead, these are probably due to amplification and scoring inconsistencies. It is possible that certain tissue types are more likely to have amplification failure than others. For example, some types of leaf tissue could have higher concentrations of PCR-inhibiting secondary metabolites. Given our result that basal-cambium samples from the same tree had higher genetic distances on average than other tissue type pairs, we wondered if samples from cambium tissue were more prone to null alleles due to lower template DNA concentrations in the PCRs. However, when we looked at the allelic phenotypes of samples in cambium-basal pairs with nonzero genetic distances, we found that only four cambium samples were missing peaks that were present in the corresponding basal sample, whereas seven basal samples were missing peaks that were present in the matching cambium sample.

Another explanation for the greater genetic distances between basal-cambium pairs could be that, unlike paired leaf tissues, basal and cambium samples were always run on separate PCR plates. Our analysis showed that amplification on different plates did cause slightly greater genetic distances between samples. This result underscores the importance of optimizing PCRs for different primers. It also provides an argument for randomizing the order of samples during DNA extraction and amplification to prevent bias due to the grouping of samples collected in close geographic proximity. In this study, the effect of amplification differences between plates was not enough to cause genotyping inaccuracy, as duplicate sample pairs consistently had genetic distances below our threshold of 0.2 and our positive control sample had a consistent genotype in all runs.

In terms of detecting differences in somatic mutation rates between tissue types, our results were inconclusive. We only detected microsatellite repeat regions that seemed to vary in length between duplicate samples in RWDI11, a marker that amplified dinucleotide repeats. In this marker, the only shifting in length of microsatellite repeats occurred where several different alleles were only one base pair apart. Rather than somatic mutation, we believe that single base pair differences in the size of microsatellite repeat regions in samples from the same tree were due to slight error in the measurement of DNA size fragments with respect to size standards. If we had seen alleles in duplicate samples shifting up or down by one repeat length in other markers as well, this would have been stronger evidence for somatic mutation.

While our microsatellite data from different tissue types from the same tree allowed us to verify the effectiveness of our genotyping protocol, it was not ideal for measuring rates of somatic mutation. Because microsatellite data only provide information on fragment length, and null alleles are often present, it was impossible to distinguish between somatic mutation and scoring error. Single-nucleotide polymorphism (SNP) or sequence data, where single base-pair changes in the genome can be detected, would be a more appropriate way to test for somatic mutations between tissue types. It would also be useful to conduct a study using all four tissue types (foliage, epicormic sprouts, basal sprouts, and cambium) from every tree sampled, which was not possible at our collection locations.

One approach not used in this study is to sample megagametophyte tissue, which would have allowed us to look at the maternal haplotype contributing to zygotes. Sampling of megagametophyte tissue may have the potential to improve allele frequency estimates, because the megagametophytes of coast redwood should be triploid, rather than hexaploid. However, triploid megagametophytes would still have some allele copy number ambiguity, making this approach less useful than it might be in a tetraploid organism. While it is possible to separate megagametophyte tissue from embryo tissue in redwood seeds (Rogers, 1997), we chose not use megagametophytes for the development of a clonal identification protocol for coast redwood. Scoring haploid tissues instead of the full hexaploid genome could reduce the power of our microsatellite markers, and issues caused by copy number ambiguity would remain. Due to the immense height of coast redwood trees, it would be very difficult to get seeds from every tree in a 1-ha plot, particularly

if the exact locations of genotypes were desired. Although analyzing megagametophyte tissue did not seem like a viable option for genotyping coast redwood, it may be an extremely useful tool in parentage and population genetic studies of this species.

In summary, a combination of optimizing PCRs, developing a conservative allele scoring protocol, and allowing for nonzero genetic distances in clonal identification allowed us to effectively identify multilocus lineages from multiple types of coast redwood tissue. We confirmed the effectiveness of our protocol using simulations and paired samples from the same trees. The techniques described in this paper will allow us to accurately identify coast redwood clones from available tissue types and have broad applicability to genetic studies of polyploid organisms, particularly where multiple tissue types are being sampled.

LITERATURE CITED

Arnaud-Haond, S., M. Migliaccio, E. Diaz-Almela, S. Teixeira, M. S. van de Vliet, F. Alberto, G. Procaccinim, et al. 2007a. Vicariance patterns in the Mediterranean Sea: East-west cleavage and low dispersal in the endemic seagrass *Posidonia oceanica*. *Journal of Biogeography* 34: 963–976.

Arnaud-Haond, S., C. M. Duarte, F. Alberto, and E. A. Serrão. 2007b. Standardizing methods to address clonality in population studies. *Molecular Ecology* 16: 5115–5139.

Bruno, D., and C. Brinegar. 2004. Microsatellite markers in coast redwood (*Sequoia sempervirens*). *Molecular Ecology Notes* 4: 482–484.

Bruvo, R., N. K. Michiels, T. G. D'Souza, and H. Schulenburg. 2004. A simple method for the calculation of microsatellite genotype distances irrespective of ploidy level. *Molecular Ecology* 13: 2101–2106.

Clark, L. V., and M. Jasieniuk. 2011. POLYSAT: An R package for polyploidy microsatellite analysis. *Molecular Ecology Resources* 11: 562–566.

Cullings, K. W. 1992. Design and testing of a plant-specific PCR primer for ecological and evolutionary studies. *Molecular Ecology* 1: 233–240.

Del Tredici, P. 1998. Lignotubers in *Sequoia sempervirens*: Development and ecological significance. *Madrono* 45: 255–260.

Douhovnikoff, V., A. M. Cheng, and R. S. Dodd. 2004. Incidence, size and spatial structure of clones in second-growth stands of coast redwood, *Sequoia sempervirens* (Cupressaceae). *American Journal of Botany* 91: 1140–1146.

Douhovnikoff, V., and R. S. Dodd. 2011. Lineage divergence in coast redwood (*Sequoia sempervirens*), detected by a new set of nuclear microsatellite loci. *American Midland Naturalist* 165: 22–37.

Dufresne, F., M. Stift, R. Vergilino, and B. K. Mables. 2014. Recent progress and challenges in population genetics of polyploid organisms: An overview of current state-of-the-art molecular and statistical tools. *Molecular Ecology* 23: 40–69.

Esselink, G. D., H. Nybom, and B. Vosman. 2004. Assignment of allelic configuration in polyploids using the MAC-PR (microsatellite DNA allele counting-peak ratios) method. *Theoretical and Applied Genetics* 109: 402–408.

Lorimer, C. G., D. J. Porter, M. A. Madej, J. D. Stuart, S. D. Veirs Jr., S. P. Norman, K. L. O'Hara, and W. J. Libby. 2009. Presettlement and modern disturbance regimes in coast redwood forests: Implications for the conservation of old-growth stands. *Forest Ecology and Management* 258: 1038–1054.

Lynch, M. 1990. The similarity index and DNA fingerprinting. *Molecular Biology and Evolution* 7: 478–484.

Neal, R. L. Jr. 1967. Sprouting of old-growth redwood stumps: First year after logging. USDA Forest Service, Pacific Southwest Research Station Research Note PSW-137. U.S. Department of Agriculture, Forest Service, Pacific Southwest Forest and Range Experiment Station, Berkeley, California.

Oliver, C. D., and B. C. Larson. 1996. Forest stand dynamics, updated edition. John Wiley & Sons, New York, New York, USA.

Parks, J. C., and C. R. Werth. 1993. A study of spatial features of clones in a population of bracken fern, *Pteridium aquilinum*. *American Journal of Botany* 80: 537–544.

Pfeiffer, T., A. M. Roschanski, J. R. Pannell, G. Korbecka, and M. Schnittler. 2011. Characterization of microsatellite loci and reliable genotyping in a polyploid plant, *Mercurialis perennis* (Euphorbiaceae). *Journal of Heredity* 102: 479–488.

R Core Team. 2014. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. Website http://www.R-project.org/ [accessed 10 February 2015].

Ramage, B. S., K. L. O'Hara, and B. T. Caldwell. 2010. The role of fire in the competitive dynamics of coast redwood forests. *Ecosphere* 1: art20. doi:10.1890/ES10-00134.1.

Rogers, D. L. 1997. Inheritance of allozymes from seed tissues of the hexaploid gymnosperm, *Sequoia sempervirens* (D. Don) Endl. (Coast redwood). *Heredity* 78: 166–175.

Rogers, D. L. 2000. Genotypic diversity and clone size in old-growth populations of coast redwood (*Sequoia sempervirens*). *Canadian Journal of Botany* 78: 1408–1419.

Rogers, D. L., and R. D. Westfall. 2007. Spatial genetic patterns in four old-growth populations of coast redwood. *In* R. B. Standiford, G. A. Guisti, Y. Valachovic, W. J. Zielinski, and M. J. Furniss [eds.], Proceedings of the Redwood Region Forest Science Symposium: What does the future hold?, 59–63. USDA Forest Service General Technical Report PSW-GTR-194. USDA Forest Service, Pacific Southwest Research Station, Albany, California, USA.